Indag. Mathem., N.S., 14 (1), 135–147 March 31, 2003

Fine and Wilf words for any periods

by R. Tijdeman and L. Zamboni

Leiden University, Mathematical Institute, Postbus 9512, 2300 RA Leiden, The Netherlands
e-mail: tijdeman@math.leidenuniv.nl
Department of Mathematics, PO Box 311430, University of North Texas, Denton, TX 76203-1430, USA
e-mail: luca@unt.edu

Communicated at the meeting of January 27, 2003

ABSTRACT

Let $w = w_1 \dots w_n$ be a word of maximal length n, and with a maximal number of distinct letters for this length, such that w has periods $p_1, \dots, p_r$ but not period $\gcd(p_1, \dots, p_r)$. We provide a fast algorithm to compute n and w. We show that w is uniquely determined apart from isomorphism and that it is a palindrome. Furthermore we give lower and upper bounds for n as explicit functions of $p_1, \dots, p_r$. For $r = 2$ the exact value of n is due to Fine and Wilf. In case the number of distinct letters in the extremal word equals r a formula for n had been given by Castelli, Mignosi and Restivo in case $r = 3$ and by Justin if $r > 3$.

1. INTRODUCTION

Let $p_1, \dots, p_r$ be positive integers. Let $w = w_1 \dots w_n$ be a word with periods $p_1, \dots, p_r$. This means that $w_{i+p} = w_i$ for $i = 1, \dots, n - p$ and $p \in \{p_1, \dots, p_r\}$. Suppose that w does not have period $\gcd(p_1, \dots, p_r)$. In 1965 Fine and Wilf [FW] showed that in case $r = 2$ the maximal length n of w equals $p_1 + p_2 - \gcd(p_1, p_2) - 1$. In 1994 de Luca and Mignosi [LM] showed that the set of all factors of non-constant words with coprime periods p_1, p_2 and length $p_1 + p_2 - 2$ coincides with the set of factors of Sturmian words. They further showed that such words w of maximal length are palindromes. In 1999 Castelli, Mignosi and Restivo [CRM] studied the case $r = 3$. They defined some function $f(x, y, z)$ such that if a word w has periods p_1, p_2, p_3 and length $\geq f(p_1, p_2, p_3)$,

---

2000 Mathematics Subject Classification 68R15, 11B50 then w has period $\gcd(p_1,p_2,p_3)$. They further showed that under suitable conditions their bound $f(x,y,z)$ is the best possible. Their construction yields maximal non-constant words with periods p_1,p_2,p_3 if and only if there is such a word of maximal length with three distinct letters. They showed that the set of factors of these words of maximal length coincides with the set of factors of all Arnoux-Rauzy sequences. Arnoux-Rauzy sequences were introduced in [AR]. Justin [J] generalized the results of Castelli, Mignosi and Restivo to words with more than three periods. For another variation of the Fine and Wilf theorem, see [MSW]. The theorem of Fine and Wilf has applications in the theory on combinatorics on words [L1], [L2], see also [T]. A multi-dimensional generalization has recently been given by Simpson and Tijdeman [ST]. For earlier work in this direction see Amir and Benson [AB], Galil and Park [GP], Giancarlo and Mignosi [GM], and Mignosi, Restivo and Silva [MRS].

Notice that the periods $p_1,...,p_r$ only induce relations between letters at places i and j when i and j are in the same residue class modulo $\gcd(p_1,...,p_r)$. It is therefore no restriction to consider such a residue class. Hence we can assume without loss of generality that $\gcd(p_1,...,p_r) = 1$. We shall do so in the sequel. If the maximal length of w under the gcd-condition is n, then the maximal length in the general case equals $(n+1)\gcd(p_1,...,p_r) - 1$. (See Section 4, Remark 2.) In particular, for proving the Fine and Wilf theorem stated above it suffices to prove that the maximal length of w equals p_1+p_2-2 provided that $\gcd(p_1,p_2) = 1$.

In Section 2 we present the algorithm to compute the extreme n and w for given periods $p_1,...,p_r$ subject to $\gcd(p_1,...,p_r) = 1$. It is interesting that the corresponding multi-dimensional continued fraction also occurs in the study of ergodic properties of a dynamical system arising from percolation theory, see [KM]. In Section 3 we present some examples. In Section 4 we give proofs of our claims and derive some properties of the extreme word w. In particular we show that it is a palindrome. In Section 5 we indicate how to compute the frequencies of the letters in w without constructing w. In Section 6 we derive lower and upper bounds for n as explicit functions of $p_1,...,p_r$.

2. THE ALGORITHM

Let positive integers $p_1,...,p_r$ be given with $\gcd(p_1,...,p_r) = 1$. We present an algorithm to construct the non-constant word $w = w_1...w_n$ with coprime periods $p_1,...,p_r$ of maximal length n and among such words with a maximal number of distinct letters occurring in w. Apart from isomorphism this word is uniquely determined, as will be shown in Theorem 5. We refer to this word as the Fine and Wilf word for periods $p_1,...,p_r$.

We first illustrate the algorithm by an example. We construct the Fine and Wilf word for periods $p_1 = 127$, $p_2 = 189$, $p_3 = 222$, $p_4 = 235$, $p_5 = 243$, $p_6 = 248$. We initialize by putting $p_1[0] := p_1, ..., p_r[0] := p_r$; $k := 0$, $m[0] := 0$. In Table 1 there are columns for $p_1[k], ..., p_r[k]$, k, $g[k]$ and $m[k]$ where $k = 0,1,2,...,K$. For each k we define $g[k+1]$ as the smallest column number which realizes the positive minimum value of $p_j[k]$. We underline the number $p_{g[k+1]}[k]$ for given k. Further we subtract $p_{g[k+1]}[k]$ from all positive periods except for the period in column $g[k+1]$ itself. If a 0 occurs in a period column, no further entries are defined in that column. We add $p_{g[k+1]}[k]$ to $m[k]$ to obtain $m[k+1]$. We repeat this so-called Reduction step until we arrive at a level K where for no i with $g[k]=i$ for some $k \leq K$ we have $p_i[K] > 1$. Finally we introduce an extra column denoting $n[k] := n - m[k]$. This column can only be completed after the value of $n (:= m[K])$ has been established. The number n is the length of the wanted Fine and Wilf word.

Example 1

We construct the extremal word for periods

$$p_1 = 127, p_2 = 189, p_3 = 222, p_4 = 235, p_5 = 243, p_6 = 248.$$

$p_1[k]$	$p_2[k]$	$p_3[k]$	$p_4[k]$	$p_5[k]$	$p_6[k]$	k	$g[k]$	$m[k]$	$n[k]$
$\underline{127}$	189	222	235	243	248	0		0	254
127	$\underline{62}$	95	108	116	121	1	1	127	127
65	62	$\underline{33}$	46	54	59	2	2	189	65
32	29	33	$\underline{13}$	21	26	3	3	222	32
19	16	20	13	$\underline{8}$	13	4	4	235	19
11	8	12	$\underline{5}$	8	5	5	5	243	11
6	$\underline{3}$	7	5	3	0	6	4	248	6
3	3	4	$\underline{2}$	0		7	2	251	3
$\underline{1}$	1	2	2			8	4	253	1
1	0	1	1			9	1	254	0

Table 1: Result of the Reduction procedure

Next we construct the Fine and Wilf word by the so-called Extension procedure as follows. By | we indicate the stage reached after an extension step. The symbols | are not part of the word. We start with the column number of the lowest 1 which is underlined. At every step the suffix of length equal to the previous underlined number is repeated if possible. However, if the underlined number exceeds the length of the constructed word, then a new letter is introduced and the complete word is repeated. The newly introduced letter is the column number of the corresponding underlined number. In both cases the number of added letters at level k equals the underlined number $p_{g[k+1]}[k]$, that is $m[k+1] - m[k] = n[k] - n[k+1]$. Therefore the number of letters which have been constructed in the Extension procedure after step k equals $n[k]$ for $k = K-1, K-2, \ldots, 0$, respectively. Thus the resulting Fine and Wilf word w has length $n[0] = m[K] = n$.

|1|41|141|41141|14141141|4114114141141|3
1 41 141 41141 14141141 4114114141141|
141 41141 14141141 4114114141141 3
1 41 141 41141 14141141 4114114141141|
1 41 141 41141 14141141 4114114141141 3
1 41 141 41141 14141141 4114114141141
141 41141 14141141 4114114141141 3
1 41 141 41141 14141141 4114114141141|
Result w of the Extension procedure.

We now formalize the construction. We initialize by putting

$$p_1[0] := p_1, ..., p_r[0] := p_r; k := 0, m[0] := 0.$$

Reduction step

(R1) determine the smallest i with $p_i[k] = \min\{p_j[k] | p_j[k] \neq 0; j = 1, \ldots, r\}$.
(R2) for $j = 1, \ldots, r$ with $p_j[k] = p_i[k]$ and $j > i$, put $p_j[k+1] := 0$;
(R3) for $j = 1, \ldots, r$ with $p_j[k] > p_i[k]$ put $p_j[k+1] := p_j[k] - p_i[k]$;
(R4) put $k := k + 1, p_i[k] := p_i[k-1], g[k] := i, m[k] := m[k-1] + p_i$.

It will turn out that the maximal word length n is equal to the lastly found value $m[k]$. We denote the last found value of k by K, whence $n = m[K]$. For the construction of the wanted word w of length n we initialize by putting $j := 1$, $s := 1$ and maintaining the values found in the Reduction procedure. Then we repeat the following Extension step until k reaches the value 0:

Extension step

(E1) If $s \geq j$, then put $w_j := g[k], w_i := w_{i-j}$ for $i = j + 1, \ldots, 2j - 1$, and put $j := 2j$,
otherwise put $w_i := w_{i-s}$ for $i = j, \ldots, j + s - 1$, and put $j := j + s$;
(E2) $k := k - 1, s := m[k] - m[k-1]$.

The resulting word w is the wanted word as we shall show in Section 4. It contains r distinct letters if and only if both $p_1[K] = \ldots = p_r[K] = 1$ and $p_j \leq m[K] = n$ for $j = 1, \ldots, r$. This is exactly the case studied by Castelli, Mignosi, Restivo [CMR] and Justin [J]. In general the word w will contain the letter $j \in \{1, \ldots, r\}$ if and only if $p_j[K] = 1$ and $p_j \leq n$.

3. FURTHER EXAMPLES

Example 2. We construct the maximal word for periods

$$p_1 = 164, p_2 = 211, p_3 = 214, p_4 = 111.$$

$p_1[k]$	$p_2[k]$	$p_3[k]$	$p_4[k]$	k	$g[k]$	$m[k]$	$n[k]$
164	211	214	$\underline{111}$	0		0	257
$\underline{53}$	100	103	111	1	4	111	146
53	$\underline{47}$	50	58	2	1	164	93
6	47	$\underline{3}$	11	3	2	211	46
$\underline{3}$	44	3	8	4	3	214	43
$\underline{3}$	41	0	5	5	1	217	40
3	38		$\underline{2}$	6	1	220	37
$\underline{1}$	36		2	7	4	222	35
$\underline{1}$	35		1	8	1	223	34
$\underline{1}$	34		0	9	1	224	33
..	..			..	..	..	..
1	1			42	1	257	0

Table 2: Result of the reduction procedure.

We have omitted a number of trivial rows, since we could foresee what the last row would be. In practice we can stop the procedure as soon as some p attains the value 1. The value of $n[k]$ at that row k will be $\max_j p_j[k] - 1$ where the maximum is taken over all j for which the column of $p_j[k]$ contains an underlined number. The column $n[k]$ can be computed from there upwards.

Example 3. It may happen that periods are induced by the other periods and therefore irrelevant. They are never activated by underlining. The columns which have no underlined numbers should be neglected when calculating $n[k]$ and in particular $n = n[K]$.

$p_1[k]$	$p_2[k]$	$p_3[k]$	$p_4[k]$	$p_5[k]$	$p_6[k]$	$p_7[k]$	k	$g[k]$	$m[k]$	$n[k]$
$\underline{9}$	16	25	14	22	19	18	0		0	19
9	7	16	$\underline{5}$	13	10	9	1	1	9	10
4	$\underline{2}$	11	5	8	5	4	2	4	14	5
$\underline{2}$	2	9	3	6	3	2	3	2	16	3
2	0	7	$\underline{1}$	4	1	0	4	1	18	1
1		6	1	3	0		5	4	19	0

Table 3: Example with irrelevantly long periods.

The extreme word is $|4|14|14|41414|141441414|$. This is also the Fine and Wilf word for periods 9, 16 and 14 (the columns with an underlined entry). This word has induced period 18 smaller than n which is characteristic for columns p_j which do not contain underlined entries and for which $p_j[K]$ is not defined, and further obviously periods 25, 22, 19 each at least n which is characteristic for columns p_j which do not contain underlined entries and for which $p_j[K]$ is defined.

Examples 1.3 (continued). We write Table 1 with the rows in inverse order. We number the rows not from $k = K$ to $k = 0$, but from $l = 0$ to $l = K$. We write $p_j(l), g(l), n(l)$ for $p_j[K-l], g[K-l], n[K-l]$, respectively. In the new table we underline the numbers $p_{g(l)}(l)$. They were overlined in Table 1.

$p_1(l)$	$p_2(l)$	$p_3(l)$	$p_4(l)$	$p_5(l)$	$p_6(l)$	l	$g(l)$	$n(l)$
$\underline{1}$	0	1	1			0	1	0
1	1	2	$\underline{2}$			1	4	1
3	$\underline{3}$	4	2	0		2	2	3
6	3	7	$\underline{5}$	3	0	3	4	6
11	8	12	5	$\underline{8}$	5	4	5	11
19	16	20	$\underline{13}$	8	13	5	4	19
32	29	$\underline{33}$	13	21	26	6	3	32
65	$\underline{62}$	33	46	54	59	7	2	65
$\underline{127}$	62	95	108	116	121	8	1	127
127	189	222	235	243	248	9		254

Table 4: Inverse table of Table 1.

In order to construct Table 4 it suffices to know the row for $l = 0$, the column for $g(l)$ and the places of the zeros. The other entries follow from:

$$p_{g(l)}(l+1) = p_{g(l)}(l),$$
$$p_j(l+1) = p_j(l) + p_{g(l)}(l) \text{ for } j \neq g(l)$$

when $p_j(l)$ is a defined integer, and

$$n(l+1) = n(l) + p_{g(l)}(l) \text{ for } l = 0, 1, \ldots, K-1.$$

For Table 3 the inverse table reads as follows.

$p_1(l)$	$p_2(l)$	$p_3(l)$	$p_4(l)$	$p_5(l)$	$p_6(l)$	$p_7(l)$	l	$g(l)$	$n(l)$
1		6	$\underline{1}$	3	0		0	4	0
$\underline{2}$	0	7	1	4	1	0	1	1	1
2	$\underline{2}$	9	3	6	3	2	2	2	3
4	2	11	$\underline{5}$	8	5	4	3	4	5
$\underline{9}$	7	16	5	13	10	9	4	1	10
9	16	25	14	22	19	18	5		19

Table 5: Inverse table of Table 3.

4. PROOFS

We start with some lemmas which we shall use repeatedly. The first result is an extension of Lemma 2.1 of [CMR].

Lemma 1. *Let $v = w_1 \ldots w_n$ be a word with s distinct letters and periods $p_1 < \ldots < p_r$. If $n \geq 2p_1$, then $u := w_1 \ldots w_{n-p_1}$ is a word with s distinct letters and periods $p_1, p_2 - p_1, \ldots, p_r - p_1$. If $n = 2p_1 - 1$, then $u := w_1 \ldots w_{n-p_1}$ is a word with at least $s - 1$ distinct letters and periods $p_1, p_2 - p_1, \ldots, p_r - p_1$.*

Proof. Because of the period p_1, every letter of w occurs in v unless $n = 2p_1 - 1$ and letter w_{p_1} does not occur elsewhere in w. In the exceptional case u has $s - 1$ distinct letters. For $m \leq n - p_j$ we have $w_m = w_{m+p_j} = w_{m+p_j-p_1}$. So u has period $p_j - p_1$ for $j = 2, \ldots, r$. $\square$

Lemma 2. *Suppose $v = w_1 \ldots w_n$ has periods $p_1, \ldots, p_r$ and $n \geq p_1$. Put $w_{n+i} = w_{n+i-p_1}$ for $i = 1, \ldots, p_1$. Then the word $w := w_1 \ldots w_{n+p_1}$ has periods $p_1, p_2 + p_1, \ldots, p_r + p_1$.*

Proof. It is obvious that the new word has period p_1. For $m \leq n - p_j$ we have $w_m = w_{m+p_j} = w_{m+p_j+p_1}$ for $j = 2, \ldots, r$. $\square$

Lemma 3. *Let $v = w_1 \ldots w_n$ be a palindrome and for some $m < n$ let $u := w_1 \ldots w_m$ be a palindrome too. Then $w := w_1 \ldots w_n w_{m+1} w_{m+2} \ldots w_n$ is also a palindrome.*

Proof. If $x = x_1 \ldots x_k$, then we write $\overline{x} = x_k \ldots x_1$ for the reverse word. We have $u = \overline{u}$ and $v = \overline{v}$. Put $t = w_{m+1} \ldots w_n$. Then

$$\begin{aligned} \overline{w} &= \overline{vt} = \overline{t}\,\overline{v} = \overline{t}v = \overline{t}ut \\ &= \overline{t}\,\overline{u}\,t = (\overline{ut})t = \overline{v}t = vt = w. \quad \square \end{aligned}$$

The next two lemmas refer to the Extension procedure. We refer to Tables 4 and 5 for illustration. By $p_j(l)$ we indicate the value of p_j at level l, if it is defined. At level $l = 0$ we start with at least two values $p_j(0)$ equal to 1, others undefined or nonnegative integers. We further put $n(0) = 0$. For $l = 0, 1, 2, \ldots, K - 1$ we select a column number $i = g(l)$ for which $p_i(l)$ is defined and $0 < p_i(l) \leq n(l) + 1$. (In the table this is indicated by underlining $p_i(l)$). Put $p_i(l+1) := p_i(l), p_j(l+1) := p_j(l) + p_i(l)$ for each $j \neq i$ for which $p_j(l)$ has been defined, and $n(l+1) := n(l) + p_i(l)$. We further have the option to define $p_j(l+1) := 0$ for some values j for which $p_j(l)$ is not defined. We call a column j activated at level k if $g(k) = j$ or $p_j(k) = 0$. We say that column j is active at level l if there exists a $k < l$ such that column j is activated at level k.

Lemma 4. *If a column j is active at level l, then $p_j(l) \leq n(l)$. Otherwise $p_j(l) = n(l) + p_j(0)$.*

Proof. If column j is not active at level l and $p_j(0) > 0$, then $p_j(k) - p_j(k-1) = n(k) - n(k-1)$ for $k = 1, \ldots, l$, whence $p_j(l) = p_j(0) + n(l) - n(0) = n(l) + p_j(0)$. If column j is active at level l and it was activated for the last time at level k, then $n(k+1) \geq p_{g(k)}(k) = p_j(k+1)$ and $p_j(h) - p_j(h-1) = n(h) - n(h-1)$ for $h = k+2, \ldots, l$ so that $n(l) \geq p_j(l)$. $\square$

Lemma 5. *If $g(l) = i$, then $n(l+1) \geq 2p_i(l) - 1$. Equality holds if and only if column i is activated for the first time at level l and $p_i(0) = 1$.*

Proof. Suppose $g(l) = i$. We have $n(l+1) = n(l) + p_i(l)$. If column i is active at level l, then $p_i(l) \leq n(l)$ by Lemma 4, whence $n(l+1) = n(l) + p_i(l) \geq 2p_i(l)$. If column i is inactive at level l, then $p_i(l) > n(l)$. Since $i = g(l)$, we deduce from the definition of K that $p_i(l) = n(l) + 1$. Hence $n(l+1) = n(l) + p_i(l) = 2p_i(l) - 1$. $\square$

The proofs of the theorems below are based on the observation that the Reduction procedure and the Extension procedure are each others inverses. If the Reduction procedure is started from periods $p_1 < \ldots < p_r$ and the Extension procedure is executed with the corresponding choices, then finally the periods $p_1, p_2, \ldots, p_r$ are attained. The given examples serve as illustrations for the proofs.

Theorem 1. *The word $w = w_1 \ldots w_n$ constructed according to the algorithm has periods $p_1, \ldots, p_r$.*

Proof. By induction on the level l. At level $l = 1$ the constructed word $w(1)$ consists of one letter, hence $w(1)$ has any period. Suppose at level l the constructed word $w(l)$ has period $p_j(l)$ for every column j which is active at level l. Let $g(l) = i$. Then, by the definitions of $p_j(l+1)$ and $w(l+1)$ and by Lemma 2, $w(l+1)$ has period $p_j(l+1)$ for every column which is active at level l. Suppose that column j is activated at level l. Then $p_j(l+1) = p_i(l+1) = p_i(l)$. Since $w(l+1)$ has period $p_i(l+1)$, it has also period $p_j(l+1)$. Thus $w(l+1)$ has period $p_j(l+1)$ for every column j which is active at level $l+1$. The induction step is complete. We conclude that at the end of the Extension procedure w has period p_j for every column j which is active at level K. If column j is not activated at all, then $p_j(0) > 0 = n(0)$ and $g(l) \neq j$ for every l. Hence, by $p_j(l+1) - p_j(l) = n(l+1) - n(l)$, we have $p_j(l) > n(l)$ for every l and in particular $p_j(K) > n(K)$. Thus p_j is larger than the length n of the constructed word w and therefore w has period p_j. $\square$

Theorem 2. *Let $p_1, p_2, \ldots, p_r (r > 1)$ be coprime positive integers. If the word constructed by the algorithm in Section 2 has length n, then every word w of length greater than n and periods $p_1, \ldots, p_r$ is constant.*

Proof. Suppose there is a word $v = v_1 \ldots v_{n+1}$ of length $n+1$ with $s > 1$ distinct letters and coprime periods $p_1 < \ldots < p_r$. By Lemma 1, $v[1] := v_1 \ldots v_{n+1-p_1}$ has periods $p_1, p_2 - p_1, \ldots, p_r - p_1$. By Lemma 5,

$$n = n[0] = n(K) \geq \min 2p_{g(K-1)}(K-1) - 1 = 2p_1 - 1.$$

Hence the length of v is at least $2p_1$ and therefore, by Lemma 1, $v[1]$ contains s distinct letters too. Thus $v[1]$ is a word with s distinct letters of length $n[1] + 1$ and with periods $p_1[1], p_2[1], \ldots, p_r[1]$. By applying Lemmas 1 and 5 inductively on k we find that $v[k] := v_1 \ldots v_{n[k]+1}$ has periods $p_1[k], p_2[k], \ldots, p_r[k]$ in so far defined and contains s distinct letters. Since at the final level K there are at least two periods 1 and no periods larger than 1 in activated columns, the underlined period at level $K-1$ equals 1. Furthermore, by the definition of K, we have $n(1) = n[K-1] = n - m[K-1] = m[K] - m[K-1] = p_{g[K]} = 1$. So we reach the conclusion that $v[K-1] := v_1 v_2$ has period 1 and contains $s > 1$ distinct letters. This is a contradiction. $\square$

Theorem 3. *Let $p_1,\ldots,p_r (r > 1)$ be coprime positive integers. The word constructed by the algorithm in Section 2 has s distinct letters if and only if s is the number of periods 1 in activated columns at the bottom row of the Reduction procedure. Hence $s > 1$.*

Proof. If $p_j(0) = 0$ or not defined, then $p_j(l) \leq n(l)$ whenever $j = g(l)$ so that column j does not yield a new letter in the constructed word. If $p_j(0) > 0$ and column j is for the first time activated at level l, then $p_j(0) = 1$ and $p_j(l) = n(l) + 1$. Hence column j introduces a new letter j at level $l + 1$. For $k > l$ we have $p_j(k) \leq n(k)$ by Lemma 4 and other activations of column j will not introduce new letters. If $p_j(0) > 0$ and column j is never activated, then p_j is an induced period and the Fine and Wilf word is determined by the other columns. □

Theorem 4. *The word w constructed by the algorithm in Section 2 is a palindrome.*

Proof. We apply induction on l. Both $w(1) = w_1$ and $w(2) = w_1w_1$ or $w_1w_2w_1$ are palindromes. Suppose both $w(l-1) = w_1 \ldots w_{n(l-1)}$ and $w(l) = w_1 \ldots w_{n(l)}$ are palindromes. Then, by Lemma 3, $w(l+1) = w_1 \ldots w_{n(l+1)}$ is a palindrome too. Thus $w = w(K) = w_1 \ldots w_{n(K)}$ is a palindrome. □

Corollary. *Let $p_1 < p_2 < \ldots < p_r (r > 1)$ be coprime positive integers. There exists a non-constant palindrome word w with periods $p_1,\ldots,p_r$ of word length $n \geq 2p_1 - 1$. A nonconstant word with coprime periods p_1 and p_2 cannot have length greater than $p_1 + p_2 - 2$.*

Proof. The word w constructed by the algorithm in Section 2 has periods $p_1,\ldots,p_r$ by Theorem 1, is non-constant by Theorem 3, is a palindrome by Theorem 4, and has word length $n \geq 2p_{g(K-1)}(K-1) - 1 = 2\min_i p_i - 1 = 2p_1 - 1$ by Lemma 5. By the Fine and Wilf theorem applied to the coprime numbers p_1 and p_2 we have $|w| = n \leq p_1 + p_2 - 2$. □

More precise bounds for n will be presented in Section 6.

We call the words $v = v_1 \ldots v_n$ with letter set A and $w = w_1 \ldots w_n$ with letter set B isomorphic if and only if there exists a bijection $f : A \to B$ such that $f(v_i) = w_i$ for $i = 1, 2, \ldots, n$.

Theorem 5. *The Fine and Wilf word for coprime periods $p_1,\ldots,p_r (r > 1)$ is uniquely determined apart from isomorphism.*

Proof. Let n be the smallest positive integer for which coprime periods $p_1 < \ldots < p_r$ exist with two nonisomorphic Fine and Wilf words $v = v_1 \ldots v_n$ and $w = w_1 \ldots w_n$. We know from the Corollary that $n \geq 2p_1 - 1$. Consider the prefixes of v and w of length $n - p_1$.

If $n = 2p_1 - 1$, then $v = v_1 \ldots v_{n-p_1} v_0 v_1 \ldots v_{n-p_1}$ and $w = w_1 \ldots w_{n-p_1} w_0 w_1 \ldots w_{n-p_1}$ by having period p_1. Moreover, $\hat{v} := v_1 \ldots v_{n-p_1}$ and $\hat{w} := w_1 \ldots w_{n-p_1}$ have coprime periods $p_1, p_2 - p_1, \ldots, p_r - p_1$ by Lemma 1. If v_0 occurs among $v_1, \ldots, v_{n-p_1}$, then we can replace v_0 with a letter $\hat{v}_0$ not occurring among $v_1, \ldots, v_{n-p_1}$ and the word $v_1 \ldots v_{n-p_1} \hat{v}_0 v_1 \ldots v_{n-p_1}$ would contain more distinct letters than v does. So v would not be a Fine and Wilf word. For the same reason the letter w_0 does not occur among $w_1, \ldots, w_{n-p_1}$. Thus $\hat{v}$ and $\hat{w}$ are non-isomorphic words of the same length $< n$ and with the same number of distinct letters. By the induction hypothesis they both differ from the Fine and Wilf word for periods $p_1, p_2 - p_1, \ldots, p_r - p_1$. Let $\hat{u}$ be the Fine and Wilf word for the periods $p_1, p_2 - p_1, \ldots, p_r - p_1$. Then, by the induction hypothesis, $\hat{u}$ is longer than $\hat{v}$, or it has more distinct letters than $\hat{v}$ has. Putting $\hat{u} = u_1 \ldots u_m$ and $u = u_1 \ldots u_m u_{m-p_1+1} u_{m-p_1+2} \ldots u_m$ we obtain a word with coprime periods $p_1 < \ldots < p_r$ according to Lemma 2, which is either longer than v or has more distinct letters (choosing for u_{m-p_1+1} a letter not occurring among $u_1, \ldots, u_m$ if $m = n - p_1 = p_1 - 1$). Thus v is not a Fine and Wilf word.

If $n \geq 2p_1$, then the argument is along the same lines, but simpler. Since the nonisomorphic words $\hat{v}$ and $\hat{w}$ both have coprime periods $p_1, p_2 - p_1, \ldots, p_r - p_1$, length $n - p_1$ and an equal number of distinct letters, by the induction hypothesis, the Fine and Wilf word $\hat{u} = u_1 \ldots u_m$ for periods $p_1, p_2 - p_1, \ldots, p_r - p_1$ is either longer, or it has length $n - p_1$ and it contains more distinct letters than $\hat{v}$ does. Put $u := u_1 \ldots u_m u_{m-p_1+1} \ldots u_m$. This word has coprime periods $p_1 < \ldots < p_r$ and it is either longer than v or it contains more distinct letters than v does. This contradicts that v is a Fine and Wilf word. □

Remark 1. Different period sets can yield the same Fine and Wilf word. For example, both the period set 13, 19, 21 and the period set 13, 17, 23 yield the same Fine and Wilf word *abababababababaababababababa*. The shortest period set yielding this Fine and Wilf word is 13,15.

Remark 2. If w has periods $p_1, \ldots, p_r$ and $\gcd(p_1, \ldots, p_r) = d > 1$, then we can apply the above to each residue class mod d, so to words $w_i w_{i+d} w_{i+2d} \ldots w_{i+md}$. If we find that the maximal length of the non-constant word with periods $\frac{p_1}{d}, \ldots, \frac{p_r}{d}$ equals n_d, then the maximal length of the word with periods $p_1, \ldots, p_r$, but not with period d, equals $d(n_d + 1) - 1$. To construct such a word, we take a non-constant word of length n_d with periods $\frac{p_1}{d}, \ldots, \frac{p_r}{d}$, write the consecutive letters at places $d, 2d, \ldots, n_d d$ and write the same letter at all other places among $1, 2, \ldots, n_d d + d - 1$. This word has the required properties. Every longer word has to have period d because of the maximality of n_d.

5. FREQUENCIES

It is possible to compute the number of occurrences of each letter in the constructed word without constructing the word w. Let $p_1, p_2, \ldots, p_r$ be coprime positive integers and let w be the non-constant word w of maximal length constructed according to the algorithm in Section 2. Let $w(l)$ be the word constructed at level l. Then by construction the number $f_j(l+1)$ of occurrences of a letter j in $w(l+1)$ is either 1 (when the letter is introduced newly) or $2f_j(l)$ minus the number of occurrences of j in the prefix of w of length $n(l) - (n(l+1) - n(l)) = n(l) - p_{g(l)}(l)$. If $n(l) - p_{g(l)}(l) \leq 0$, then the frequencies of the not newly introduced letters of j are doubled. Recall that $n(i+1) - n(i) = p_{g(l)}(i+1) = p_{g(l)}(i)$ if $g(l) = g(i)$ and $n(i+1) - n(i) = p_{g(l)}(i+1) - p_{g(l)}(i)$ if $p_{g(l+1)} > 0$ and $g(l) \neq g(i)$. Hence, if $n(l) > p_{g(l)}(l)$, then $n(l) - p_{g(l)}(l) = n(i)$ where i is the largest integer less than l for which column $g(l)$ was activated.

We conclude that if $n(l+1) = 2n(l) + 1$, then $f_j(l+1) = 2f_j(l)$ for $j \neq g(l)$ and $f_{g(l)}(l+1) = 1$. If $n(l+1) \leq 2n(l)$, then $f_j(l+1) = 2f_j(l) - f_j(i)$ where i is the last level before l when column $g(l)$ was activated. In the latter case $w(i)$ has length $n(i) = 2n(l) - n(l+1)$. This provides an easy way to determine i.

Example 4. We compute the frequencies of the letters in the Fine and Wilf word for periods 158, 65, 142, 112, 156, 130.

p_1	p_2	p_3	p_4	p_5	p_6	l	$g(l)$	$n(l)$	$f(1)$	$f(4)$	$f(6)$
1			$\underline{1}$		1	0	4	0	0	0	0
$\underline{2}$			1	0	2	1	1	1	0	1	0
2			3	$\underline{2}$	4	2	5	3	1	2	0
4		0	5	2	$\underline{6}$	3	6	5	2	3	0
10		6	11	8	$\underline{6}$	4	6	11	4	6	1
16		$\underline{12}$	17	14	6	5	3	17	6	9	2
28	0	12	29	26	$\underline{18}$	6	6	29	10	15	4
46	18	30	$\underline{47}$	44	18	7	4	47	16	24	7
93	$\underline{65}$	77	47	91	65	8	2	94	32	48	14
158	65	142	112	156	130	9		159	54	81	24

Table 6: The frequencies in the word for periods 158, 65, 142, 112, 156, 130

6. ESTIMATES FOR THE LENGTH OF THE MAXIMAL WORD

We present bounds for the length n of the non-constant word w of maximal length with coprime periods $p_1 < p_2 < \ldots < p_r$. These bounds depend only on $p_1, \ldots, p_r$. They are refinements of the inequalities $2p_1 - 1 \leq n \leq p_1 + p_2 - 2$ stated in the Corollary. We assume that p_i does not divide p_j for $i \neq j$, since otherwise period p_i implies period p_j and period p_j can be neglected. We further assume $r \geq 3$.

Lower bounds. Put $t = \lfloor p_2/p_1 \rfloor$. Then in the Reduction procedure p_i is subtracted t times whereafter $p_2 - tp_1$ is the minimal value. Subsequently we subtract u times $p_2 - tp_1$ where $u = \min\lfloor \frac{p_1}{p_2 - tp_1}, \frac{p_3 - tp_1}{p_2 - tp_1} \rfloor$. The next minimal value is either $p_1 - u(p_2 - tp_1)$ or $p_3 - tp_1 - u(p_2 - tp_1)$ whichever is smaller. We now apply the lower bound $n(l+1) \geq 2p_{g(l)}(l) - 1$ from Lemma 5 in the form $n[k-1] \geq 2p_{g[k]}[k] - 1$ for $k = t + u + 1$. Hence

$$n[t+u] \geq 2(p_1 - u(p_2 - tp_1)) - 1 \text{ when } p_3 \geq (t+1)p_1$$

and

$$n[t+u] \geq 2(p_3 - up_2 + (u-1)tp_1) - 1 \text{ when } p_3 < (t+1)p_1.$$

This yields, by $n - n[t+u] = tp_1 + u(p_2 - tp_1)$, the following bounds:

(a) $\quad n \geq ((u+1)t+2)p_1 - up_2 - 1$ when $p_3 \geq (t+1)p_1$,

(b) $\quad n \geq 2p_3 - up_2 + t(u-1)p_1 - 1$ when $p_3 < (t+1)p_1$.

In Examples 1, 2 and 3 we have $t = 1, u = 1$ and we obtain the lower bounds 254, 257, 17 from (b), whereas the exact values are 254, 257, 19, respectively. In Example 5 below we have $t = 2, u = 4$ and we find $n \geq 149$ from (b) whereas n=154.

Example 5. An example with $t > 1, u > 1$.

We consider $p_1 = 50, p_2 = 111, p_3 = 147, p_4 = 149$. Note that t is the number of consecutive times that $p_1 (= 50)$ is underlined and u the number of consecutive times that $p_2[t] (= 11)$ is underlined.

$p_1[k]$	$p_2[k]$	$p_3[k]$	$p_4[k]$	k	$g[k]$	$m[k]$	$n[k]$
$\underline{50}$	111	147	149	0		0	154
$\underline{50}$	61	97	99	1	1	50	104
50	$\underline{11}$	47	49	2	1	100	54
39	$\underline{11}$	36	38	3	2	111	43
28	$\underline{11}$	25	27	4	2	122	32
17	$\underline{11}$	14	16	5	2	133	21
6	11	$\underline{3}$	5	6	2	144	10
3	8	3	$\underline{2}$	7	3	147	7
$\underline{1}$	6	1	2	8	4	149	5
$\underline{1}$	5	0	1	9	1	150	4
$\underline{1}$	4		0	10	1	151	3
$\underline{1}$	3			11	1	152	2
$\underline{1}$	2			12	1	153	1
1	1			13	1	154	0

Table 7: Reduction procedure for periods 50, 111, 147, 149.

Upper bounds. Suppose that $p_2 - tp_1, p_3 - tp_1$ are pairwise coprime. If we follow the argument above until level $l = t$, we find, by the upper bound from the Corollary,

$$n[t] \leq (p_2 - tp_1) + (p_3 - tp_1) - 2.$$

Hence, from $n - n[t] = tp_1$, we obtain

(c) $\quad n \leq p_2 + p_3 - tp_1 - 2.$

In Examples 1, 2, 3, 5 we find from (c) the upper bounds 282, 262, 19, 156, where the exact values are 254, 257, 19, 154, respectively.

REFERENCES

[AB] Amir, A. and G.E. Benson – Alphabet independent two-dimensional pattern matching. Proc. 24th ACM Symp. Theor. Comp., pp. 59–68 (1992).

[AR] Arnoux, P. and G. Rauzy – Représentation geométrique de suites de complexité $2n + 1$. Bull. Soc. Math. France **119**, 199–215 (1991).

[CMR] Castelli, M.G., F. Mignosi and A. Restivo – Fine and Wilf's theorem for three periods and a generalization of Sturmian words. Theor. Computer Sc. **218**, 83–94 (1999).

[FW] Fine, N.J. and H.S. Wilf – Uniqueness theorem for periodic functions. Proc. Amer. Math. Soc. **16**, 109–114 (1965).

[GM] Giancarlo, R. and F. Mignosi – Generalizations of the periodicity theorem of Fine and Wilf. CAAP '94, LNCS 787, Springer, Berlin etc., pp. 130–141 (1994).

[GP] Galil, Z. and K. Park – Truly alphabet independent two-dimensional pattern matching. Proc. 33th IEEE Symp. on Foundations of Computer Science, pp. 247–256 (1992).

[J] Justin, J. – On a paper by Castelli, Mignosi, Restivo. RAIRO: Theoret. Informatics Appl. **34**, 373–377 (2000).

[KM] Kraaikamp, C. and R. Meester – Ergodic properties of a dynamical system arising from percolation theory. Ergod. Th. Dynam. Sys. **15**, 653–661 (1995).

[L1] Lothaire, M. (ed.) – Combinatorics on Words. Addison Wesley, Reading MA (1983).

[L2] Lothaire, M. (ed.) – Algebraic Combinatorics on Words. Cambridge University Press (2002).

[LM] Luca, A. de and F. Mignosi – Some combinatorial properties of finite words. Theor. Computer Sc. **136**, 361–385 (1994).

[MRS] Mignosi, F., A. Restivo and P.V. Silva – On Fine and Wilf's theorem for bidimensional words. Theor. Computer Sc. **292**, 245–262 (2003).

[MSW] Mignosi, F., J. Shallit and M.-w. Wang – Variations on a theorem of Fine and Wilf. MFCS 2001, LNCS 2136, Springer, Berlin etc., pp. 512–523 (2001).

[ST] Simpson, R.J. and R. Tijdeman – Multi-dimensional versions of a theorem of Fine and Wilf and a formula of Sylvester. Proc. Amer. Math. Soc. **131**, 1661–1671 (2003.)

[T] Tijdeman, R. – Intertwinings of periodic functions. Indag. Math. N.S. **9**, 113–122 (1998).